Proceedings of the 8th International Conference on Structural Dynamics, EURODYN 2011
Leuven, Belgium, 4-6 July 2011
G. De Roeck, G. Degrande, G. Lombaert, G. Müller (eds.)
ISBN 978-90-760-1931-4

Bayesian updating, model class selection and robust stochastic predictions of structural response

James L.Beck[1]
[1]Departments of Mechanical and Civil Engineering, and Computing and Mathematical Sciences,
California Institute of Technology, Pasadena, CA, USA
email: jimbeck@caltech.edu

ABSTRACT: A fundamental issue when predicting structural response by using mathematical models is how to treat *both* modeling and excitation uncertainty. A general framework for this is presented which uses probability as a multi-valued conditional logic for quantitative plausible reasoning in the presence of uncertainty due to incomplete information. The fundamental probability models that represent the structure's uncertain behavior are specified by the choice of a stochastic system model class: a set of input-output probability models for the structure and a prior probability distribution over this set that quantifies the relative plausibility of each model. A model class can be constructed from a parameterized deterministic structural model by stochastic embedding utilizing Jaynes' Principle of Maximum Information Entropy. Robust predictive analyses use the entire model class with the probabilistic predictions of each model being weighted by its prior probability, or if structural response data is available, by its posterior probability from Bayes' Theorem for the model class. Additional robustness to modeling uncertainty comes from combining the robust predictions of each model class in a set of competing candidates weighted by the prior or posterior probability of the model class, the latter being computed from Bayes' Theorem. This higher-level application of Bayes' Theorem automatically applies a quantitative Ockham razor that penalizes the data-fit of more complex model classes that extract more information from the data. Robust predictive analyses involve integrals over high-dimensional spaces that usually must be evaluated numerically. Published applications have used Laplace's method of asymptotic approximation or Markov Chain Monte Carlo algorithms.

KEY WORDS: Structural modeling; Robust stochastic analysis; System identification; Bayesian updating; Ockham's razor.

1 INTRODUCTION

A common practice during dynamic design of a structure, or design of a response control system for a structure, is to use a single mathematical model to predict its dynamic response to prescribed wind or seismic excitations. Often this model is developed using finite-element software. The structural model predictions, on their own, are not very useful, however, unless they give information about their accuracy. The response predictions will have uncertain accuracy not only because of the uncertainty in the future structural excitations but also because the structural model will always involve approximations of the real dynamic behavior that produce uncertain affects in the predicted response; in addition, the structural model will usually involve parameters whose values are uncertain. This structural modeling uncertainty, in addition to future excitation uncertainty, should be explicitly treated when making predictive analyses.

In the case of an existing structure where response sensor data is available, the modeling uncertainty can be reduced by updating the mathematical model of the structure, thereby allowing more accurate predictions of its future response to specified excitations. This process is commonly called *system identification* and, as usually practiced, it consists primarily of *parameter estimation* where data from the structure is used to estimate the value of the uncertain parameter vector specifying the model. Parameter estimation may be done in several ways: by maximizing the posterior PDF (*probability density function*) from Bayes' Theorem to get the MAP (*maximum a posteriori*) estimate; by maximizing the likelihood function to get the MLE (*maximum likelihood estimate*), which is equivalent to the MAP estimate under a uniform prior over the parameter space (or some sub-region of it); or by LS (*least-squares*) output matching, which is equivalent to the MLE under a joint Gaussian probability model for the combined prediction and measurement errors, which are defined as the difference between the measured and model outputs.

There are several conceptual and computational difficulties with parameter estimation:

1) Since any structural model can only be expected to approximate the real structural behavior, there are no true values of the parameters. Therefore, determining a single "best" value for the parameter vector, such as its LS estimate or MLE, and using that model to make response predictions, is a questionable procedure;

2) The parameter estimate (ML or LS) is often not unique, especially for complex multi-parameter structural models, and so then there are multiple models with multiple corresponding response predictions. The common procedure of arbitrarily fixing some of the parameter values so that the remaining ones can be uniquely estimated may severely bias the predictions. A more principled approach should be taken that includes all of the multiple predictions in an appropriate way;

3) No model of the system is expected to give perfect predictions so it is important to explicitly quantify the uncertain prediction errors.

The tasks of explicitly quantifying modeling and excitation uncertainty in response predictions during design and operation can be done in a rigorous probabilistic framework. The theory for treating excitation uncertainty, known as *random vibrations* or more commonly nowadays, as *stochastic dynamics* (or *mechanics*), has a long history. On the other hand, the theory and computational tools for a probabilistic treatment of modeling uncertainty are more recent because their development was hampered by the commonly-taught restrictive interpretation of probability as the relative frequency of "inherently random" events in the "long run", which does not provide a meaning for the probability of a parameter value or a model.

In this paper, we describe a general stochastic (i.e. probabilistic) framework for handling both modeling and excitation uncertainty when predicting the dynamic response of a structure based on structural models. It uses an interpretation of probability as a logic for quantitative plausible reasoning when there is uncertainty due to incomplete information. The foundations of probability logic are due to the physicists Cox [1,2] and Jaynes [3,4]. The vague and speculative concept of inherent randomness is not needed. We consider both *prior robust stochastic analysis*, which is appropriate during structural design, and *posterior robust stochastic analysis*, which can be performed for an existing structure if response sensor data is available. Before giving an overview of the theory for these robust predictive analyses, we first provide a brief summary of probability logic and then we define a stochastic system model class which provides the required fundamental probability models for robust stochastic analyses.

2 PROBABILITY LOGIC

In probability logic, probability is viewed as a multi-valued conditional logic for plausible reasoning that extends binary Boolean propositional logic to the case of incomplete information. The probability $\mathrm{P}[b|c]$ is interpreted as the degree of plausibility of the proposition (statement) b based on the information in the proposition c where c is only conditionally asserted. This interpretation is consistent with the Bayesian perspective that probability represents a degree of belief in a proposition.

For a propositional calculus of plausible reasoning involving probabilities, we need to evaluate the following probabilities in terms of more basic ones: $\mathrm{P}[\sim b|c]$, $\mathrm{P}[a \; \& \; b|c]$ and $\mathrm{P}[a \textit{ or } b|c]$, which correspond, respectively, to the degree of plausibility based on c that b is not true, that both a and b are true, and that either a or b (or both) are true. Cox [1] *derived* the appropriate calculus by extending the axioms of Boolean logic which deals with the special case of complete information where the truth or falsity of b is known from c, that is, $\mathrm{P}[b|c]=1$ or $\mathrm{P}[b|c]=0$, respectively.

Cox's results can be stated as a minimal set of axioms for probability logic. For any propositions a, b, c:

(i) $\mathrm{P}[b|c] \geq 0$
(ii) $\mathrm{P}[\sim b|c] = 1 - \mathrm{P}[b|c]$ (Negation Function)
(iii) $\mathrm{P}[a\&b|c] = \mathrm{P}[a|b\&c]\mathrm{P}[b|c]$ (Conjunction Function)

Using the last two axioms and De Morgan's Law from Boolean logic, we can derive:

$\mathrm{P}[a \text{ or } b|c] = \mathrm{P}[a|c]+\mathrm{P}[b|c] - \mathrm{P}[a \; \& \; b|c]$ (Disjunction Function)

The axioms for a probability measure $P(A)$ on subsets A of a finite set X, as stated by Kolmogorov [5] and commonly given in textbooks on probability theory, can be derived as a special case of the probability logic axioms where the propositions refer to uncertain membership of an object in a set [6]. For example, if X denotes the set of possible values for an uncertain-valued variable x, then for any subset A of X, $P(A)$ can be interpreted as $\mathrm{P}[x \in A \mid \pi]$ where π denotes the proposition that states $x \in X$ and specifies the probability model for x quantifying the relative degree of plausibility of each value of x in X. Kolmogorov also defines conditional probability in terms of unconditional probabilities but in probability logic all probabilities are inherently conditional and so the corresponding result appears as an axiom (Conjunction Function).

The probability logic axioms therefore provide a calculus for handling variables whose values are uncertain. The vague and speculative concept of inherent randomness is not needed so the axioms can not only be applied to variables that correspond to physical quantities but also to models and model parameters, in contrast to the relative frequency interpretation of Kolmogorov's axioms. This allows robust probabilistic predictions that account for modeling uncertainty.

3 STOCHASTIC SYSTEM MODEL CLASSES

3.1 Definition of a model class

In modeling the I/O (input-output) behavior of a system, one cannot expect any chosen deterministic model to make perfect predictions and the prediction errors of any such model will be uncertain. This motivates the introduction of a *stochastic system* (or Bayesian) *model class* $\mathcal{M}$ that consists of fundamental probability models to describe the uncertain I/O behavior of the system: a set of I/O probability models $\{p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M}): \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{N_p}\}$ and a prior probability model $p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$ that expresses the initial probability of each model $p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M})$, that is, the prior gives a measure of the initial relative plausibility of the I/O probability models corresponding to each value of the parameter vector $\boldsymbol{\theta}$. Here, $\mathbf{u}$ and $\mathbf{x}$ denote the system input and output vectors that consist of discretized time histories of the excitation and corresponding system response.

The probability models defining the model class $\mathcal{M}$ are viewed as representing a state of knowledge about the structural system conditional on the available information and not as its inherent properties. All probabilistic predictions for the structure are conditional on the chosen fundamental probability models for the model class, which we make explicit in the notation by conditioning on $\mathcal{M}$.

3.2 Model class construction by stochastic embedding

Any deterministic dynamic model of a structural system that involves uncertain parameters (e.g. a finite-element structural model) can be used to construct a model class $\mathcal{M}$ for the system by *stochastic embedding* [7]. Suppose that the deterministic model defines an implicit or explicit mathematical relationship $\mathbf{q}(\mathbf{u},\boldsymbol{\theta})$ between the input $\mathbf{u}$ and model output $\mathbf{q}$ where both are discretized time histories and the uncertain model parameters are denoted by $\boldsymbol{\theta}$. The first step is to introduce the uncertain prediction-error time history $\mathbf{e}$ [8] as the difference between the *real system output* $\mathbf{x}$ and the *model output* $\mathbf{q}$ for the same input, i.e. $\mathbf{x} = \mathbf{q} + \mathbf{e}$, so $\mathbf{e}$ provides a bridge between the model world and the real world.

The next step is to establish a parameterized probability model for $\mathbf{e}$ by using the Principle of Maximum Information Entropy [4], which states that the probability model should be selected to produce the most uncertainty (largest Shannon entropy) subject to parameterized constraints that we wish to impose; the selection of any other probability model would lead to an unjustified reduction in the amount of prediction uncertainty. The maximum-entropy probability model is therefore conservative in the sense that it gives the greatest uncertainty in the prediction-error time history, and hence in the system-output time history, conditional on what one is willing to assert about the system.

A simple choice for the probability model for $\mathbf{e}$ is produced by choosing the following constraints during entropy maximization: zero prediction-error mean at each time (any uncertain bias can be added to $\mathbf{q}$ as another uncertain parameter), and a parameterized prediction-error variance or covariance matrix at each time. The maximum entropy PDF for the prediction error $\mathbf{e}$ over an unrestricted range is then discrete-time Gaussian white noise. Therefore, the I/O probability model for the system output $x_i \in \mathbb{R}^{N_o}$ at discrete time t_i, conditional on the parameter vector $\boldsymbol{\theta}$, is given by the following Gaussian PDF with the mean equal to the model output $q_i(\mathbf{u},\boldsymbol{\theta}) \in \mathbb{R}^{N_o}$ and with a parameterized covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \in \mathbb{R}^{N_o \times N_o}$:

$$p(x_i|\,\mathbf{u},\boldsymbol{\theta},\mathcal{M}) = \frac{1}{(2\pi)^{N_o/2}\,|\Sigma|^{1/2}} \exp\left[-\tfrac{1}{2}(x_i - q_i)^T \Sigma^{-1}(x_i - q_i)\right]$$

The I/O probability model for the system output history $\mathbf{x}$ over N discrete times is then given by:

$$p(\mathbf{x}|\,\mathbf{u},\boldsymbol{\theta},\mathcal{M}) = \Pi_{i=1}^{N}\; p(x_i|\,\mathbf{u},\boldsymbol{\theta},\mathcal{M})$$

The stochastic independence exhibited here comes from the fact that no joint moments in time are imposed as constraints during the entropy maximization. It refers to information independence which is not necessarily causal independence. It asserts that if the prediction errors at certain discrete times are given, this does not influence the plausibility of the prediction-error values at other times.

Another choice for stochastic embedding [7,9] is to use a state-space model of the structure and introduce prediction errors into the state vector equation, as well as in the output equation, again modeled with a Gaussian PDF based on the Principle of Maximum Information Entropy. This alternative allows updating of the prediction-error uncertainty at unobserved points in the system, not just at the measurement points.

Either choice for the stochastic modeling of the prediction errors produces a set of parameterized I/O probability models $\{p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M}) : \boldsymbol{\theta} \in \Theta\}$ where the uncertain parameters $\boldsymbol{\theta}$ now also include those involved in specifying the probability models for the prediction errors, such as the prediction-error variances. To complete the specification of the model class $\mathcal{M}$, a prior distribution $p(\boldsymbol{\theta}|\mathcal{M})$ is chosen to express the relative plausibility of each I/O probability model $p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M})$ specified by the parameter vector $\boldsymbol{\theta}$.

3.3 *Bayesian updating within a model class*

Suppose system data $\mathcal{D} = \{\hat{\mathbf{u}}, \hat{\mathbf{x}}\}$ is available that consists of measured output $\hat{\mathbf{x}}$ of the system and possibly the corresponding system input $\hat{\mathbf{u}}$. These data can be used to update the relative plausibility of each I/O probability model $p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M})$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^{N_P}$, in the set defined by a model class $\mathcal{M}$ by computing the *posterior* PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ from Bayes' Theorem:

$$p(\boldsymbol{\theta}\,|\,\mathcal{D},\mathcal{M}) = c^{-1} p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M})\,p(\boldsymbol{\theta}\,|\,\mathcal{M}) \tag{1}$$

where $c = p(\mathcal{D}\,|\mathcal{M}) = \int_\Theta p(\mathcal{D}\,|\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$ is the normalizing constant, and $p(\mathcal{D}\,|\boldsymbol{\theta},\mathcal{M})$, as a function of $\boldsymbol{\theta}$, is the *likelihood function* which expresses the probability of getting data $\mathcal{D}$ based on the PDF $p(\mathbf{x}|\mathbf{u},\boldsymbol{\theta},\mathcal{M})$ for the system output. The constant $c = p(\mathcal{D}\,|\mathcal{M})$ is also called the *evidence* for the model class $\mathcal{M}$ given by data $\mathcal{D}$. Although it is a normalizing constant in (1) and so it does not affect the shape of the posterior distribution, it plays an important role in computing the posterior probability of the model class, as described later. The likelihood function should strictly be denoted by $p(\hat{\mathbf{x}}\,|\,\hat{\mathbf{u}},\boldsymbol{\theta},\mathcal{M})$ but the notation used in (1) is convenient.

4 ROBUST PREDICTIVE ANALYSIS USING A MODEL CLASS

A model class can be used to perform both *prior* (initial) and *posterior* (updated using system sensor data) *robust predictive analyses* based purely on the probability axioms [10]. Prior robust analyses are of importance in the robust design of systems whereas posterior robust analyses can be used to improve predictive modeling of already operating systems.

Based on a selected model class $\mathcal{M}$, all the probabilistic information for the prediction of the discrete response time history $\mathbf{x}$ for a specified discrete input time history $\mathbf{u}$ is contained in the *prior robust predictive* PDF given by the Total Probability Theorem as:

$$p(\mathbf{x}\,|\,\mathbf{u},\mathcal{M}) = \int p(\mathbf{x}\,|\,\mathbf{u},\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}\,|\,\mathcal{M})d\boldsymbol{\theta} \tag{2}$$

Notice that (2) gives a weighted average of the probabilistic prediction $p(\mathbf{x}|\,\mathbf{u},\boldsymbol{\theta},\mathcal{M})$ for each model specified by $\boldsymbol{\theta} \in \Theta$ in model class $\mathcal{M}$ where the weight is given by the prior probability $p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$.

If system sensor data $\mathcal{D}$ is available from the structure, the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ can be computed from Bayes Theorem as in Section 3.3, then the *posterior robust predictive* PDF is given by:

$$p(\mathbf{x}\,|\,\mathbf{u},\mathcal{D},\mathcal{M}) = \int p(\mathbf{x}\,|\,\mathbf{u},\boldsymbol{\theta},\mathcal{D},\mathcal{M})p(\boldsymbol{\theta}\,|\,\mathcal{D},\mathcal{M})d\boldsymbol{\theta} \tag{3}$$

These prior and posterior robust predictions correspond to a type of integrated global sensitivity analysis where the probabilistic prediction of each I/O probability model specified by the model class is considered but it is weighted by the relative plausibility of the model according to the prior or posterior PDF, respectively, in accordance with the Total Probability Theorem.

Usually in assessing a structure's design, the response time history **x** is not directly used but instead a *system performance measure* is selected that, because of the modeling uncertainty, is expressed as the *prior* or *posterior* expected value of some *performance function* $\mathbf{g}(\mathbf{x})$:

$$\mathrm{E}[\mathbf{g}(\mathbf{x}) \mid \mathbf{u}, \mathcal{M}] = \int \mathbf{g}(\mathbf{x}) p(\mathbf{x} \mid \mathbf{u}, \mathcal{M}) d\mathbf{x} \approx \frac{1}{\mathrm{I}} \sum_{i=1}^{\mathrm{I}} \mathbf{g}(\mathbf{x}^{(i)})$$

or: (4)

$$\mathrm{E}[\mathbf{g}(\mathbf{x}) \mid \mathbf{u}, \mathcal{D}, \mathcal{M}] = \int \mathbf{g}(\mathbf{x}) p(\mathbf{x} \mid \mathbf{u}, \mathcal{D}, \mathcal{M}) d\mathbf{x} \approx \frac{1}{\mathrm{I}} \sum_{i=1}^{\mathrm{I}} \mathbf{g}(\mathbf{x}^{(i)})$$

where, as shown, the integrals can be approximated using standard MCS (Monte Carlo simulation) by drawing samples $\boldsymbol{x}^{(i)}$, i = 1,2,...,I, from $p(\mathbf{x}| \mathbf{u},\mathcal{M})$ or $p(\mathbf{x}| \mathbf{u},\mathcal{D},\mathcal{M})$.

Usually there is also uncertainty in the input **u**, for example from future wind or seismic excitation of the structure, which can be described by a stochastic model $\mathcal{U}$ that specifies a joint PDF $p(\mathbf{u}|\mathcal{U})$ for the discrete input time history **u**. This uncertainty in the excitation can then be incorporated by evaluating the additional integral:

$$\mathrm{E}[\mathbf{g}(\mathbf{x}) \mid \mathcal{U}, \mathcal{M}] = \int \mathrm{E}[\mathbf{g}(\mathbf{x}) \mid \mathbf{u}, \mathcal{M}] p(\mathbf{u} \mid \mathcal{U}) d\mathbf{u} \approx \frac{1}{\mathrm{J}} \sum_{j=1}^{\mathrm{J}} \mathrm{E}[\mathbf{g}(\mathbf{x}) \mid \mathbf{u}^{(j)}, \mathcal{M}] \qquad (5)$$

or its posterior counterpart based on (3) and (4). As shown in (5), this integral over all inputs **u** defined by $\mathcal{U}$ can be approximated using standard MCS where the theoretical mean of $\mathrm{E}[\mathbf{g}(\mathbf{x})| \mathbf{u},\mathcal{M}]$ with respect to $p(\mathbf{u}|\mathcal{U})$ in (5) is approximated by its sample mean based on J samples $\boldsymbol{u}^{(j)}$, j = 1,2,...,J, drawn from $p(\mathbf{u}|\mathcal{U})$.

An important special case is where $\mathbf{g}(\mathbf{x})=\mathrm{I}_F(\mathbf{x})$ is an indicator function which is equal to 1 if $\mathbf{x} \in F$ and 0 otherwise, where F is a region in the response space that corresponds to unsatisfactory system performance, then (5) gives the *prior robust failure probability* $\mathrm{P}(F|\mathcal{U},\mathcal{M})$ [10]. If this failure probability is very small, a more computationally efficient algorithm than MCS should be used, such as Subset Simulation based on MCMC (Markov Chain Monte Carlo) simulation [11,12].

In optimal robust stochastic design, the system performance measure in (5) serves as the objective function in the optimization over the design variables specifying each design choice; for example, the performance function $\mathbf{g}(\mathbf{x})$ could represent the structural design's life-cycle costs and include future uncertain economic losses from seismic damage over a specified time period (e.g. [13]). Using stochastic simulation to evaluate the objective function in (5) for each iteration of the design variables given by an optimization algorithm leads to a huge computational effort. A very efficient MCMC algorithm, SSO (Stochastic Subset Optimization), has been developed [13,14,15] to find a small set of near-optimal design variables containing the optimum design, rather than trying to converge onto the point estimate at much greater computational cost.

The robust predictive models in (2) and (3) require the evaluation of multi-dimensional integrals over the parameter space that cannot usually be evaluated analytically, nor evaluated numerically in a straightforward way if the number of parameters is not very small. The *prior* robust integrals in (2), (4) and (5) can usually be readily evaluated by standard Monte Carlo simulation where for (2), samples are drawn from an appropriately selected prior PDF $p(\boldsymbol{\theta}|\mathcal{M})$. For robust reliability-based design involving small failure probabilities, however, more computationally efficient algorithms such as Subset Simulation should be used.

In contrast, evaluation of the *posterior* robust integral in (3) is much more challenging because (i) evaluation of the normalizing constant c in Bayes' Theorem (1) requires a challenging high-dimensional integration over the model parameter space; and (ii) the high probability content region of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ occupies a much smaller volume in the parameter space than that of the prior PDF and this region may be quite contorted because of the correlations between the model parameters that are induced by the data $\mathcal{D}$. Useful methods to approximate these integrals are Laplace's method of asymptotic approximation and stochastic simulation methods.

4.1 Laplace's method of asymptotic approximation

Laplace's method can be used to approximate the posterior robust integral in (3) (e.g. [8,10,16]). This method requires a non-convex optimization in what is usually a high-dimensional parameter space, which is computationally challenging, especially when the model class is not globally identifiable and so there may be multiple global maximizing points. For a general system, Beck & Katafygiotis [8] define *global system identifiability*, *local system identifiability* and *system unidentifiability based on the data* in terms of whether the set of MLEs consists of a single point, discrete points or a continuum of points in the continuous-valued parameter space, respectively.

The importance of Laplace's asymptotic approximation is that it provides a justification for the common practice of parameter estimation where just one predictive model in the model class is selected, *provided* the model class is globally identifiable based on the data and the amount of data is not too small, because applied to the integral in (2), it gives [8]:

$$p(\mathbf{x} \mid \mathbf{u}, \mathcal{D}, \mathcal{M}) \approx p(\mathbf{x} \mid \mathbf{u}, \hat{\boldsymbol{\theta}}, \mathcal{D}, \mathcal{M})$$

where $\hat{\boldsymbol{\theta}}$ is the MLE or the MAP estimate for the model class based on data $\mathcal{D}$.

4.2 Markov Chain Monte Carlo simulation methods

If the stated conditions for Laplace's approximation do not apply, then robust response predictions can be made based on Markov Chain Monte Carlo methods to approximate the integral in (2), such as multi-level Metropolis-Hastings algorithms with tempering or annealing (e.g. [17,18]), Gibbs sampler (e.g. [19]), and Hybrid Monte Carlo (or Hamiltonian Markov Chain) simulation (e.g. [20]). These MCMC methods are used to draw samples from the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$, say $\boldsymbol{\theta}^{(k)}$, $k = 1,2,...,K$, and the integral in (3) is approximated by:

$$p(\mathbf{x}\,|\,\mathbf{u},\mathcal{D},\mathcal{M}) \approx \frac{1}{K}\sum_{k=1}^{K} p(\mathbf{x}\,|\,\mathbf{u},\boldsymbol{\theta}^{(k)},\mathcal{D},\mathcal{M}) \tag{6}$$

5 ROBUST PREDICTIVE ANALYSIS USING MULTIPLE MODEL CLASSES

Sometimes it may be appropriate to choose a set of competing candidate model classes to deal with the uncertainty in choosing a model class to represent the dynamic behavior of a structure. The probability logic axioms then lead naturally to *prior* and *posterior hyper-robust predictive models* that combine the predictions of all model classes in this set. These robust predictions are especially important when calculating failure probabilities because for reliable systems they tend to be very sensitive to the particular choice of model and this sensitivity is alleviated by considering the integrated robust or hyper-robust failure probabilities (e.g. [7,9]).

If $\boldsymbol{M}$ specifies a set of candidate model classes $\{\mathcal{M}_j$: j=1,2, …,$N_M\}$ that is being considered for a system, together with a prior probability distribution over this set, then all the probabilistic information for the prediction of system response $\mathbf{x}$ subject to input $\mathbf{u}$ is contained in the *prior hyper-robust predictive* PDF based on $\boldsymbol{M}$ and the Total Probability Theorem:

$$p(\mathbf{x}\,|\,\mathbf{u},\boldsymbol{M}) = \sum_{j=1}^{N_M} p(\mathbf{x}\,|\,\mathbf{u},\mathcal{M}_j)\mathrm{P}(\mathcal{M}_j\,|\,\boldsymbol{M}) \tag{7}$$

where the prior robust predictive PDF for each model class $\mathcal{M}_j$ from (2) is weighted by the chosen prior probability $\mathrm{P}(\mathcal{M}_j\,|\,\boldsymbol{M})$, which can be chosen to be $1/N_M$ if the model classes are considered equally plausible a priori.

On the other hand, if response data $\mathcal{D}$ is available from the structure, the corresponding *posterior hyper-robust predictive* PDF based on $\boldsymbol{M}$ can be computed from:

$$p(\mathbf{x}\,|\,\mathbf{u},\mathcal{D},\boldsymbol{M}) = \sum_{j=1}^{N_M} p(\mathbf{x}\,|\,\mathbf{u},\mathcal{D},\mathcal{M}_j)\mathrm{P}(\mathcal{M}_j\,|\,\mathcal{D},\boldsymbol{M}) \tag{8}$$

where the posterior robust predictive PDF for each model class $\mathcal{M}_j$ from (3) is weighted by its posterior probability $\mathrm{P}(\mathcal{M}_j\,|\,\mathcal{D},\boldsymbol{M})$ computed from Bayes' Theorem at the model class level:

$$\mathrm{P}(\mathcal{M}_j\,|\,\mathcal{D},\boldsymbol{M}) = \frac{p(\mathcal{D}\,|\,\mathcal{M}_j)\mathrm{P}(\mathcal{M}_j\,|\,\boldsymbol{M})}{p(\mathcal{D}\,|\,\boldsymbol{M})} \tag{9}$$

Here $p(\mathcal{D}\,|\,\mathcal{M}_j)$ is the *evidence* (sometimes called marginal likelihood) for $\mathcal{M}_j$ provided by the data $\mathcal{D}$, which is given by the Total Probability Theorem as:

$$p(\mathcal{D}\,|\,\mathcal{M}_j) = \int p(\mathcal{D}\,|\,\boldsymbol{\theta}_j,\mathcal{M}_j)p(\boldsymbol{\theta}_j\,|\,\mathcal{M}_j)d\boldsymbol{\theta}_j \tag{10}$$

The posterior probability of model class $\mathcal{M}_j$ in (9) is a measure, based on data $\mathcal{D}$, of its plausibility relative to $\boldsymbol{M}$, the chosen set of candidate model classes for making structural response predictions. There is no implied assumption here that one of the model classes is the 'correct' or 'true' one.

5.1 *Calculation of the data-based evidence for a model class*

The computation of the multi-dimensional integral in (10) for the evidence is nontrivial. Laplace's method of asymptotic approximation can be used when the model class is globally identifiable based on the available data $\mathcal{D}$ (e.g. [7,21]), which gives:

$$p(\mathcal{D}\,|\,\mathcal{M}_j) \approx p(\mathcal{D}\,|\,\hat{\boldsymbol{\theta}}_j,\mathcal{M}_j)p(\hat{\boldsymbol{\theta}}_j\,|\,\mathcal{M}_j)(2\pi)^{N_j/2}\det(\mathbf{H}(\hat{\boldsymbol{\theta}}_j))^{-1/2} \tag{11}$$

where N_j is the number of model parameters (the dimension of $\boldsymbol{\theta}_j$) for the model class $\mathcal{M}_j$ and $\mathbf{H}(\boldsymbol{\theta}_j)$ is the Hessian matrix of $-\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}_j,\mathcal{M}_j)$ if the parameter estimate used in (11) is the unique MLE (maximum likelihood estimate) that maximizes $\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}_j,\mathcal{M}_j)$. However, when the chosen class of models is unidentifiable based on the available data $\mathcal{D}$ so that there are multiple MLEs, only stochastic simulation methods are practical to calculate the model class evidence, such as the Markov Chain Monte Carlo methods: TMCMC [18,22] or the stationarity method in Cheung and Beck [23].

5.2 *Quantitative Ockham razor*

A comparison of the posterior probability of each model class automatically implements a quantitative version of a *Principle of Model Parsimony* or *Ockham razor* [24,25] which states qualitatively that simpler models should be preferred over more complex models that lead to only slightly better agreement with the data, although it was not completely clear how to quantify the complexity of a model. Two well-known measures for this purpose are AIC [26] and BIC [27] which trade-off a data-fit measure with a measure of complexity:

$$\mathrm{AIC}(\mathcal{M}_j\,|\,\mathcal{D}) = \ln p(\mathcal{D}\,|\,\hat{\boldsymbol{\theta}},\mathcal{M}_j) - N_j$$

$$\mathrm{BIC}(\mathcal{M}_j\,|\,\mathcal{D}) = \ln p(\mathcal{D}\,|\,\hat{\boldsymbol{\theta}},\mathcal{M}_j) - \tfrac{1}{2}N_j\ln N$$

where N is the number of data-points in the system sensor data $\mathcal{D}$ (model classes with a larger AIC or BIC are to be preferred because of the scaling chosen here). Using these simplified criteria for model assessment requires caution, however, because their penalty term for model class complexity depends only on the number of uncertain parameters N_j, while the correct penalty term, which can be deduced by taking the logarithm of the large-N asymptotic approximation of the evidence in (11), can differ greatly for two model classes with the same number of uncertain parameters [22]. Rather than using AIC and BIC to assess globally identifiable model classes, it is much better to approximate the evidence by using (11); for example, Saito and Beck [28] use this approximation to determine the data-based most probable order of ARX models for the seismic response of a high-rise building in Tokyo where AIC did not give a maximum over the model order.

A recent interesting information-theoretic result [22,29] shows that the evidence for $\mathcal{M}_j$ explicitly builds in a trade-off between a data-fit measure of the model class and an information-theoretic measure of its complexity (the *relative entropy* or *Kullback-Liebler information* of the posterior relative to the prior) which quantifies the amount of information the model class extracts from the data $\mathcal{D}$ [6]. This result gives a deeper understanding of why the quantitative Ockham razor based on the posterior probability for each model class, as given in Eq. (9), has a built-in mechanism against data over-fitting, thereby avoiding the well-known problem that occurs when a model is judged based only on its data-fit using the maximum likelihood estimates of the model parameters.

6 CONCLUSIONS

A powerful unifying framework is available for treating modeling uncertainty, along with input uncertainty, when using dynamic models to predict structural response during design or operation of a structure. This framework is a principled one that is based solely on the probability axioms and Jaynes' Principle of Maximum Information Entropy. A key concept is a stochastic system model class which defines the fundamental probability models that allow both prior and posterior robust stochastic structural analyses to be performed. Such a model class can be constructed by stochastic embedding of any deterministic model of the structure's input-output behavior. There is no invocation of inherent randomness; instead, the approach is a pragmatic one that allows plausible reasoning about structural behavior based on incomplete information.

The prior and posterior robust predictions of structural response not only incorporate *parametric uncertainty* (uncertainty about which model in a proposed set should be used to represent the structure's input-output behavior) but also *non-parametric uncertainty* due to the existence of prediction errors because of the approximate nature of any structural model.

Robust predictive analysis involves integrals that usually cannot be evaluated in a straight-forward way. Useful computational tools are Laplace's method of asymptotic approximation and various MCMC (Markov Chain Monte Carlo) algorithms. Recent applications of MCMC methods in structural dynamics include optimal robust stochastic design (e.g. [13,14,15]), calculating robust reliability [9,17], Bayesian updating of linear structural models for structural health monitoring using changes in modal parameter estimates [19], and Bayesian updating and model class assessment of unidentifiable hysteretic structural models [22], of dynamic structural models with many uncertain parameters [20] and of stochastic state-space models of a four-story test structure [7,9].

REFERENCES

[1] R.T. Cox, Probability, frequency and reasonable expectation, *American J. of Physics*, 14: 1-13, 1946.

[2] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, MD, USA, 1961.

[3] E.T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, R.D. Rosenkrantz (ed.), D. Reidel Publishing, Dordrecht, Holland, 1983.

[4] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[5] A.N. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea Publishing, New York, 1950.

[6] J.L. Beck, *Probability Logic, Information Quantification and Robust Predictive System Analysis*, Technical Report EERL 2008-05, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, California, 2008.

[7] J.L. Beck, Bayesian system identification based on probability logic, *Structural Control and Health Monitoring*, 17: 825-847, 2010.

[8] J.L. Beck and L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, *Journal of Engineering Mechanics*, 124(4): 455-461, 1998.

[9] S.H. Cheung, *Stochastic Analysis, Model and Reliability Updating of Complex Systems with Applications to Structural Dynamics*, PhD Thesis in Civil Engineering, California Institute of Technology, Pasadena, California, 2009.

[10] C. Papadimitriou, J.L. Beck and L.S. Katafygiotis, Updating robust reliability using structural test data, *Probabilistic Engineering Mechanics*, 16: 103-113, 2001.

[11] S.K. Au and J.L. Beck, Estimation of small failure probabilities in high dimensions by Subset Simulation, *Probabilistic Engineering Mechanics*, 16: 263-277, 2001.

[12] S.K. Au and J.L. Beck, Subset Simulation and its application to seismic risk based on dynamic analysis, *Journal of Engineering Mechanics*, 129: 901-917, 2003.

[13] A.A. Taflanidis and J.L. Beck, Life-cycle cost optimal design of passive dissipative devices, *Structural Safety*, 31: 508-522, 2009.

[14] A.A. Taflanidis, *Stochastic System Design and Applications to Stochastically Robust Structural Control*, PhD Thesis in Civil Engineering, California Institute of Technology, Pasadena, California, 2007.

[15] A.A. Taflanidis and J.L. Beck, An efficient framework for optimal robust stochastic system design using stochastic simulation, *Computer Methods in Applied Mechanics and Engineering*, 198: 88-101, 2008.

[16] K.V. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*, John Wiley and Sons, New York, 2010.

[17] J.L. Beck and S.K. Au, Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation, *Journal of Engineering Mechanics*, 128: 380-391, 2002.

[18] J. Ching and Y.J. Chen, Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging, *Journal of Engineering Mechanics*, 133: 816-832, 2007.

[19] J. Ching, M. Muto and J.L. Beck, Structural model updating and health monitoring with incomplete modal data using Gibbs Sampler, *Computer-Aided Civil and Infrastructure Engineering*, 21: 242-257, 2006.

[20] S.H. Cheung and J.L. Beck, Bayesian model updating using Hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, *Journal of Engineering Mechanics*, 135: 243-255, 2009.

[21] J.L. Beck and K.V. Yuen, Model selection using response measurements: a Bayesian probabilistic approach, *Journal of Engineering Mechanics*, 130(2): 192-203, 2004.

[22] M. Muto and J.L. Beck, Bayesian updating of hysteretic structural models using stochastic simulation, *Journal of Vibration and Control*. 14: 7-34, 2008.

[23] S.H. Cheung and J.L. Beck, Calculation of the posterior probability for Bayesian model class assessment and averaging from posterior samples based on dynamic system data, *Computer-Aided Civil and Infrastructure Engineering*, 25: 304-321, 2010.

[24] S.F. Gull, Bayesian inductive inference and maximum entropy, *Maximum Entropy and Bayesian Methods*, Skilling J (ed.), Kluwer, Dordrecht, Holland, 1989.

[25] D.J.C. Mackay, *Bayesian Methods for Adaptive Models*, PhD Thesis in Computation and Neural Systems, California Institute of Technology, Pasadena, California, 1992.

[26] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19: 716-723, 1974.

[27] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics*, 6: 461-464, 1978.

[28] T. Saito and J.L. Beck, Bayesian model selection for ARX models and its application to structural health monitoring, *Earthquake Engineering and Structural Dynamics*, 39: 1737-1759, 2010.

[29] J. Ching, M. Muto and J.L. Beck, Bayesian linear structural model updating using Gibbs sampler with modal data, *Proceedings International Conference on Structural Safety and Reliability*, Rome, Italy, 2005.